Tests of Significance

Diana Mindrila, Ph.D.
Phoebe Balentyne, M.Ed.

Based on Chapter 15 of The Basic Practice of Statistics (6th ed.)

Concepts:
- The Reasoning of Tests of Significance
- Stating Hypotheses
- *P*-value and Statistical Significance
- Tests for a Population Mean
- Significance from a Table

Objectives:
- Define statistical inference.
- Describe the reasoning of tests of significance.
- Describe the parts of a significance test.
- State hypotheses.
- Define *P*-value and statistical significance.
- Conduct and interpret a significance test for the mean of a Normal population.
- Determine significance from a table.

References:
Moore, D. S., Notz, W. I, & Flinger, M. A. (2013). *The basic practice of statistics* (6th ed.). New York, NY: W. H. Freeman and Company.

Statistical Inference

Confidence intervals are one of the two most common types of statistical inference. Researchers use a confidence interval when their goal is to estimate a population parameter. The second common type of inference, called a **test of significance**, has a different goal: to assess the evidence provided by data about some claim concerning a population.

A **test of significance** is a formal procedure for comparing observed data with a claim (also called a hypothesis), the truth of which is being assessed.
- The claim is a statement about a parameter, like the population proportion p or the population mean μ.
- The results of a significance test are expressed in terms of a probability that measures how well the data and the claim agree.

The Reasoning of Tests of Significance

It is helpful to start with an example:

Population Mean:
IQ=100

Is the sample mean significantly different than the population mean?

Distribution of IQ scores

Population mean IQ=100

55   70   85   100   115   130   145

IQ Score
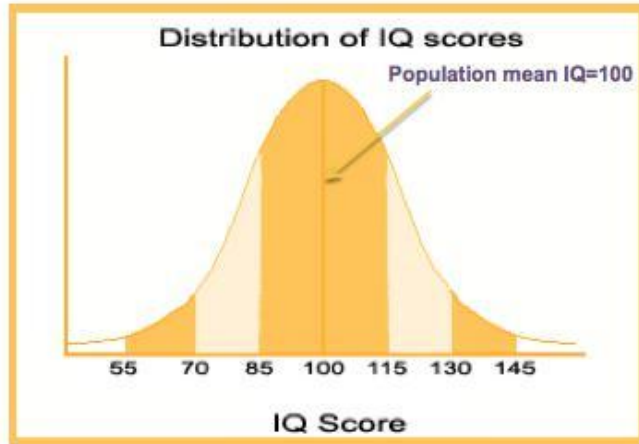
- In order to determine if two numbers are *significantly different*, a statistical test must be conducted to provide evidence. Researchers cannot rely on subjective interpretations.
- Researchers must collect statistical evidence to make a claim, and this is done by conducting a test of statistical significance.

Stating Hypotheses

The first step in conducting a test of statistical significance is to state the hypothesis.

> A significance test starts with a careful statement of the claims being compared.
>
> The claim tested by a statistical test is called the **null hypothesis (H_0)**. The test is designed to assess the strength of the evidence *against* the null hypothesis. Often the null hypothesis is a statement of "no difference."
>
> The claim about the population that evidence is being sought for is the **alternative hypothesis (H_a)**.
> The alternative is **one-sided** if it states that a parameter is *larger* or *smaller than* the null hypothesis value.
> It is **two-sided** if it states that the parameter is *different from* the null value (it could be either smaller or larger).

- When using logical reasoning, it is much easier to demonstrate that a statement is false, than to demonstrate that it is true. This is because proving something false only requires one counterexample. Proving something true, however, requires proving the statement is true in every possible situation.
- For this reason, when conducting a test of significance, a **null hypothesis** is used. The term **null** is used because this hypothesis assumes that there is no difference between the two means or that the recorded difference is not significant. The notation that is typically used for the null hypothesis is H_0.
- The opposite of a null hypothesis is called the **alternative hypothesis**. The alternative hypothesis is the claim that researchers are actually trying to prove is true. However, they prove it is true by proving that the null hypothesis is false. If the null hypothesis is false, then its opposite, the alternative hypothesis, must be true. The notation that is typically used for the alternative hypothesis is H_a.

Example:

Null Hypothesis:

Ho: IQ sample mean=100, Or IQ sample mean - 100 =0

In words:

Ho: The sample mean IQ is not significantly different than 100.

Alternative Hypothesis:

One-Sided Ha: IQ sample mean>100, Or Ha: IQ sample mean - 100 > 0

In words:

One-sided Ha: The sample mean IQ is significantly larger than 100.

Two-sided Ha: IQ>≠100 or IQ-100≠0

- In the example above, the null hypotheses states: "the sample mean is equal to 100" or "there is no difference between the sample mean and the population mean."
  - ➢ The sample mean will not be exactly equal to the population mean. This null hypothesis is stating that the recorded difference is not a significant one.
  - ➢ If researchers can demonstrate that this null hypothesis is false, then its opposite, the alternative hypothesis, must be true.
- In the example above, the alternative hypothesis states: "the sample mean *is* significantly different than 100" or "there *is* a significant difference between the sample mean and the population mean."
- If researchers are trying to prove that the mean IQ in the sample will specifically be *higher* or *lower* (just one direction) than the population mean, this is a **one-sided** alternative hypothesis because they are only looking at one direction in which the mean may vary. They are not interested in the other direction.
- If researchers suspect that the sample mean could be either lower *or* higher than 100, the alternative hypothesis would be **two-sided** because both directions in which mean IQ may vary are being tested.

- ❖ When conducting a significance test, the goal is to provide evidence to reject the null hypothesis. If the evidence is strong enough to reject the null hypothesis, then the alternative hypothesis can automatically be accepted. However, if the evidence is not strong enough, researchers fail to reject the null hypothesis.
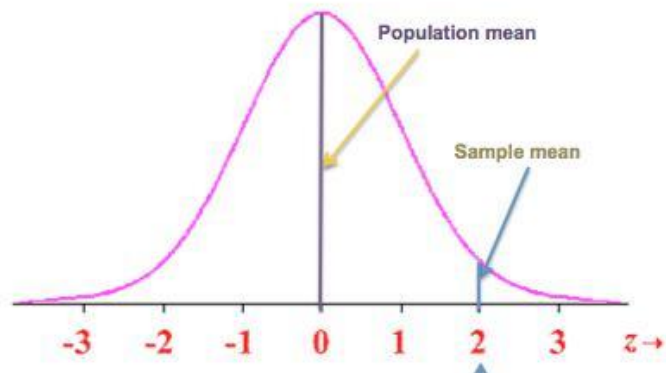
Test-Statistic (z score)

Example:

Population mean: IQ=100
Population st dev=16
Sample mean: IQ=108
Sample size: N=16

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Test Statistic

z=(Sample Mean-Population Mean)/(st deviation in the population/√Sample size)

z=(108-100)/(16/√16) => z=8/(16/4) => z=8/4 => z=2

- In the above example, the mean IQ score for the sample is 108. This is slightly higher than the population mean, which is 100. The sample mean is obviously different from the population mean, but tests of significance must be done to determine if the difference is statistically significant. The difference could possibly be attributed to chance or to sampling error.
- The first step is to compute the **test statistic**. The test statistic is simply the z score for the sample mean. The only difference is that the population standard deviation is divided by the square root of *N*, just like when a confidence interval is computed.
- To compute the test statistic, the population standard deviation must be known for the variable. The population standard deviation for IQ is 16.
- To compute the test statistic, the sample size must also be known. In this case, it is 16. (In a real research scenario, the sample size would be larger. Small sample sizes are being use in this example to make calculations simpler).
- After putting the needed information into the formula, the result is a z score of 2. This means that the sample mean is exactly two standard deviations above the population mean.
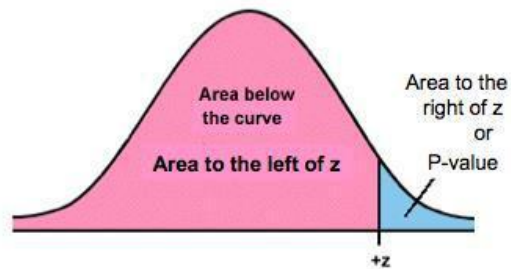
P-Value

- After computing the test statistic, the next step is to find out the probability of obtaining this score when the null hypothesis is true.
- The Normal curve helps researchers determine the percentage of individuals in the population who are located within certain intervals or above or below a certain score.
- To find this information, the score needs to be standardized. In the case of the example, this was already done by computing z, the test statistic.

The Normal distribution can be used to compute the probability of obtaining a certain z score.

Assuming that H_0 is true:

Area to the left of z = the probability of obtaining scores lower than z

Area to the right of z (p-value) = the probability of obtaining scores higher than z

The smaller the p-value, the stronger the evidence against H_0 provided by the data.

- When testing hypotheses, researchers are interested in the area to the right of z, which is called the p-value.
- The p-value represents the probability of obtaining scores that are at the z level or higher when the null hypothesis is true. In other words, what percent chance exists of getting this specific sample mean score if it is actually no different from the population mean.
- If z is far away from the mean, the p-value is small. The larger the test statistic (the farther from the mean), the smaller the p-value.
- When the p-value is very small, researchers can say they have strong evidence that the null hypothesis is false. This is because if the p-value is very small, it means that the probability of obtaining a score that is so extreme or even higher is very small.
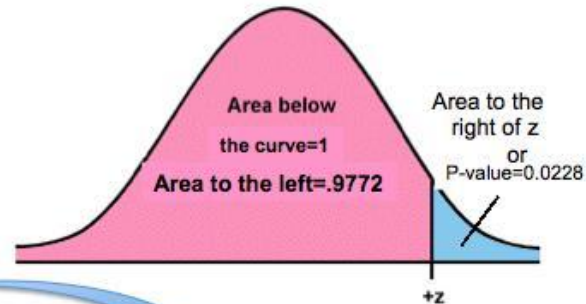
P-Value

Now it is important to know how to find the p-value.

- Statistical textbooks provide tables listing the area to the left of the curve for every possible z score.
    - ➢ The rows of this table represent the first two digits of the z score.
    - ➢ The columns of this table represent the last two digits of the z score.

Example:

Table A provides the area to the left of z:

Z	.0	.01	.02
1.9	.9713	.9719	.9726
2.0	.9772	.9778	.9783
2.1	.9821	.9826	.9830

z=2.00 => Area to the left=0.9772

P_value=1-0.9772 => P_value=0.0228

=> "If the population mean IQ is equal to 100, there is a 2% probability of recording a sample mean of 108.

Area below the curve=1
Area to the left=.9772

Area to the right of z
or
P-value=0.0228

+z

- The z score in the example is exactly 2, so all decimals are zero. The area to the left of the curve for this z score is 0.9772.
- However, for hypothesis testing, the area to the *right* of z is needed. This is called the p-value. Since the entire area under the curve is equal to one, simply subtract the area to the left of the value from one to obtain the p-value.
- In this example, 1 – 0.9772 = 0.0228.
- This means that there is only a 2% chance that the null hypothesis is true. In other words, if the population mean is 100, then there is only a 2% chance of having a sample mean equal to 108.

_P_-Value and Statistical Significance

It is important to know how small the p-value needs to be in order to reject the
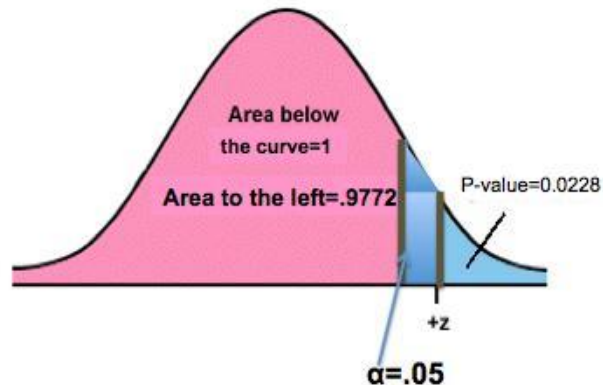null hypothesis.

**How small should the P_value be
to reject the null hypothesis?**

α (alpha) – (significance level: the probability
of rejecting Ho when Ho is true)

Is the P-value smaller than α ?

**P-value > α => accept the null
hypothesis**

**P-value <= α => reject the null
hypothesis & accept the alternative
hypothesis**

Area below
the curve=1

Area to the left=.9772

P-value=0.0228

+z

α=.05

- The cutoff value for p is called **alpha**, or the significance level.
- The researcher establishes the value of alpha prior to beginning
  the statistical analysis.
- In social sciences, alpha is typically set at 0.05 (or 5%). This represents the
  amount of acceptable error, or the probability of rejecting a null hypothesis
  that is in fact true. It is also called the probability of Type I error.

Once the alpha level has been selected and the p-value has been computed:
- ❖ If the p-value is larger than alpha, accept the null hypothesis and reject
  the alternative hypothesis.
- ❖ If the p-value is smaller than alpha, reject the null hypothesis and accept the
  alternative hypothesis.

Example:

p-value<α => Decision: Reject the null hypothesis & accept the alternative

α =.05
p-value=0.02

Conclusion: "The test of significance provided evidence that the sample IQ is
significantly larger than 100."

- In the above example, the p-value of approximately 0.02 is smaller than 0.05,
  so the null hypothesis can be rejected.
- The test of significance showed that the difference between the sample mean
  and the population mean is statistically significant.

Two-Sided Alternative Hypotheses

- A two-sided alternative hypothesis is used when there is no reason to believe that the sample mean can only be higher or lower than a given value. Researchers are only hypothesizing that the values are significantly different.
- In the example, the alternative hypothesis was one-sided, so it was only necessary to look at the probability that the sample mean was larger than 100.
- However, when the alternative hypothesis is two-sided, the sample mean can be higher or lower than the given value, so researchers must look for both extremely high and extremely low values. This means that alpha is located at both ends of the curve.
- Half of alpha is located at the higher end, and half is located at the lower end. So, there is both a low cutoff value and a high cutoff value.
- Therefore, in two-sided cases, the p-value is obtained by multiplying the area to the right of z by 2.
- Only after doubling this value can the p-value be compared to alpha.

When the alternative hypothesis is two sided:

P-value=Area to the right of the curve (from Table A)*2

Example:
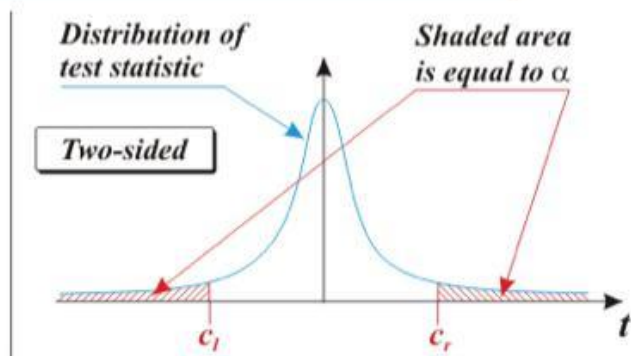Null Hypothesis: Sample IQ=100
Alternative Hypothesis: Sample IQ≠100

Z=2
Area to the right=.0228
P_Value=.0228*2= .0456

Distribution of test statistic

Shaded area is equal to α

Two-sided

c_l c_r t

Example:

$\alpha =.05$
p-value=0.04

p-value<α => Decision: Reject the null hypothesis & accept the alternative

Conclusion: "The test of significance provided evidence that the sample IQ is significantly different than 100."

- In the above example, the area to the right of z was 0.0228. So, if the alternative hypothesis were two-sided, the p-value would be 0.0456.
- The p-value (0.0456) is smaller than the alpha level (0.05), so the null hypothesis is rejected and the alternative hypothesis is accepted.

How to Obtain the *P*-Value for the z Statistic

1) A table in a statistical textbook

2) Statistical software

3) A calculator (these can be found online)

Tests of Significance: The Four-Step Process

1. **State the null and alternative hypotheses.**

2. **Calculate the test statistic.**

3. **Find the *P*-value (using a table or statistical software).**

4. **Compare P-value with α and decide whether the null hypothesis should be rejected or accepted.**

❖ The z statistic is computed based on the properties of the Normal distribution, so it should only be used when the distribution of the variable is approximately normal. If this condition is not met, a different statistic must be used.